# A contribution to the analysis of ALMA Mars observations simultaneously by 12-m and 7-m arrays

Miro Barta, EU ARC – Czech node, Astronomical Institute Ondřejov, barta@asu.cas.cz

## Introduction

This text aims to contribute to understanding, why we have so much different flux density vs. uv distance for 12-m and 7-m simultaneous observations of Mars. In particular, it addresses the discrepancy between the observed and simulated uv flux density for 12-m array as found in the analysis by Hideo Sagawa.

I think, the first point – the fact that we have somewhat different uv flux densities for 7-m and 12-m arrays – is clear now: Simply, the Mars disc does not fit completely to the 12-m primary beam, so we are integrating the brightness from the lesser angular area (NB: One could, on the other hand, expect that we have a stronger signal from bigger antenna, but the plotted signal is flux *density*, so it is already scaled-down by the antenna collecting area). The remaining problem is, why this difference is so unexpectedly large – the simulations by Hideo show that the jump between the 7m and 12m arrays should not be so big and that the expected (simulated) 12m uv power is by factor ~1.3 higher than that observed.

The latter issue (the missing factor 1.3 between observed and simulated 12m data) has been found by studying the ratio $F_{sim}$(uv_dist)/$F_{obs}$(uv_dist) in a broad range of the uv distances. In this plot by Hideo, there is another remarkable feature in addition to the fact that the average/median value of the ratio is not equal to 1.0: It is the slope of the function in between the null-points of the uv flux density – see the following Fig. 1 (taken from Hideo's presentation), where the slope is emphasized by the green line:

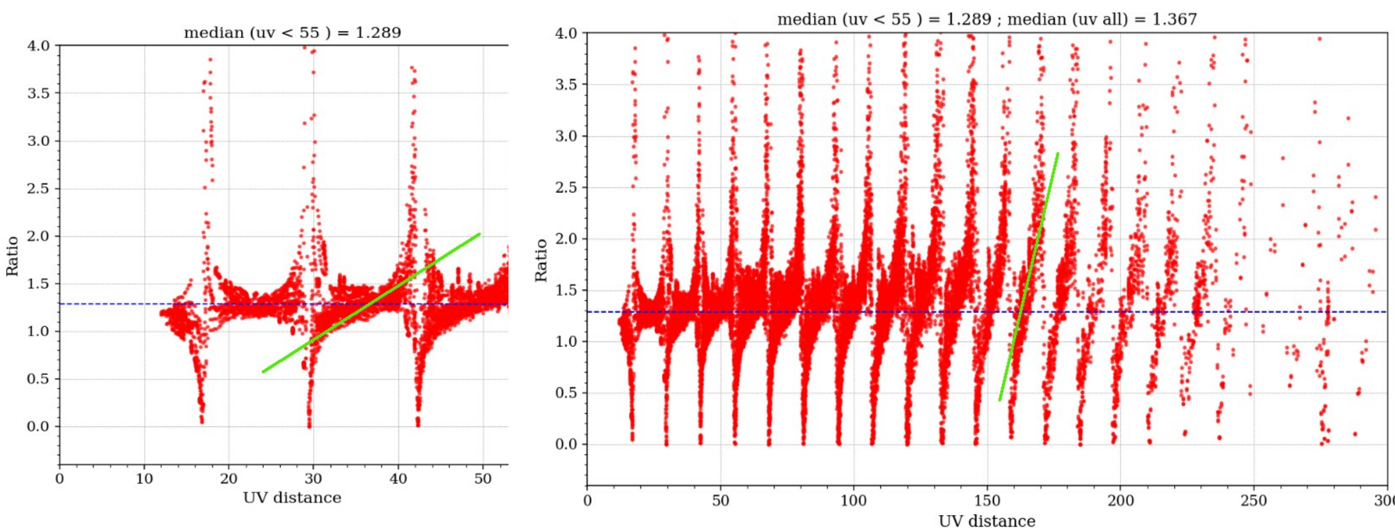


*Fig. 1: Sim vs. obs ratio. Green lines mark the unexpected slope.*

Normally, one would expect that we have by-parts constant (and equal to 1.0) ratio between the null points interrupted by vertical lines close to the null-points (because of indefinite limit of the type '0/0' and presence of noise in the observed data). Actually, exactly this can be seen in Fig. 4 further. Where does this slope come from? I think it is because the interferometric structure (the flux density vs uv-distance plot) differ for simulated and modelled 12-m data, namely the null points of those two functions differ – as marked on the next figure, again taken from the Hideo's analysis – see the following Fig 2:

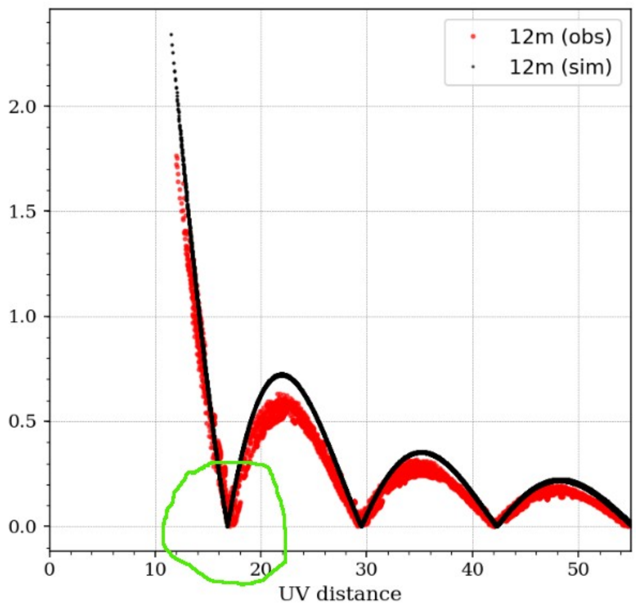

*Fig. 2: flux density vs uv distance - a discrepancy between null-points for the obs and sim data.*

As one can see, the structures are somewhat displaced/scaled in the UV-distance axis. Just a forward note: The structures (namely position of the null points) of the *simulated* data for 7m and 12m arrays differ, too. But in view that the structures depend on both the Mars disc size and antenna diameter – see below – this difference is not a mystery.

To sum up, we have actually two, likely closely related questions: (1) Why the simulated flux is (in average) by a factor ~1.3 higher than the observed one, and (2), why the null points – and the entire interferometric structure (i.e., the flux density vs uv-distance plot) – differ between the simulated and the observed data, what, consequently, results into the positive slopes in the sim/obs flux density ratio.

## So which parameters define the interferometric structure?

In order to answer the question I did a simple analytical calculation in 1D – this means like we had just a right-ascension/hour-angle '$\theta$' on the sky and corresponding Fourier component '$u$'. The reasons, why in 1D: (1) I am not brave/capable enough to calculate integrals containing the Bessel/Airy functions in their kernel, while in 1D case, with the function $\sin^2(x)/x^2$ (see relation (6) below) it is, relatively, simple. (2) The main features – like the shape of the interferometric pattern and dependence of its null-point positions on the antenna diameter and the size of the observed object (Mars) – can be *qualitatively* reproduced in the simple 1D model (up to some factor less than 1.5). And, (3) because of the simplistic assumptions: (i) ideal parabolic dish of diameter $D$ and, (ii) homogeneously bright disc of angular size $M$ as a model for the Mars, we cannot approach the real situation fully in a *quantitative* manner, anyway, even with the 2D model.

The 1D calculation basically follows the mathematical derivation of the van Zittert-Cernike theorem that stands in the foundation of interferometry. I did it by pencil on the paper, the text here summaries just the main points. The situation is sketched in the Fig. 3 (next page).

The (complex) voltage $U_1(t)$ at *Antenna 1* caused by incident radiation with the electric-field intensity $E(\upsilon,t)$ can be written as

$$\hat{U}_1(t)=K_1\int_{source}\hat{E}(\vartheta,t-\tau_D)\int_{\frac{-D_1}{2}}^{\frac{+D_1}{2}}\exp[-i\omega(t-\frac{x\vartheta}{c}-\tau_D)]\,dx\,d\vartheta \quad ,$$

which can be re-arranged (using notorious relation between wavelength $\lambda$ and angular frequency $\omega$) as

$$\hat{U}_1(t)=K_1\int_{source}\hat{E}(\vartheta,t-\tau_D)e^{-i\omega(t-\tau_D)}\int_{\frac{-D_1}{2}}^{\frac{+D_1}{2}}\exp\left(\frac{2\pi i\vartheta x}{\lambda}\right)dx\,d\vartheta$$

Here, integration over $x$ goes over the (1D ideal parabolic) antenna with diameter $D_1$, integration over the hour-angle $\upsilon$, measured as an offset to the phase-reference point $\theta_0$ (i.e., $\theta=\theta_0+\upsilon$) goes over the (1D)

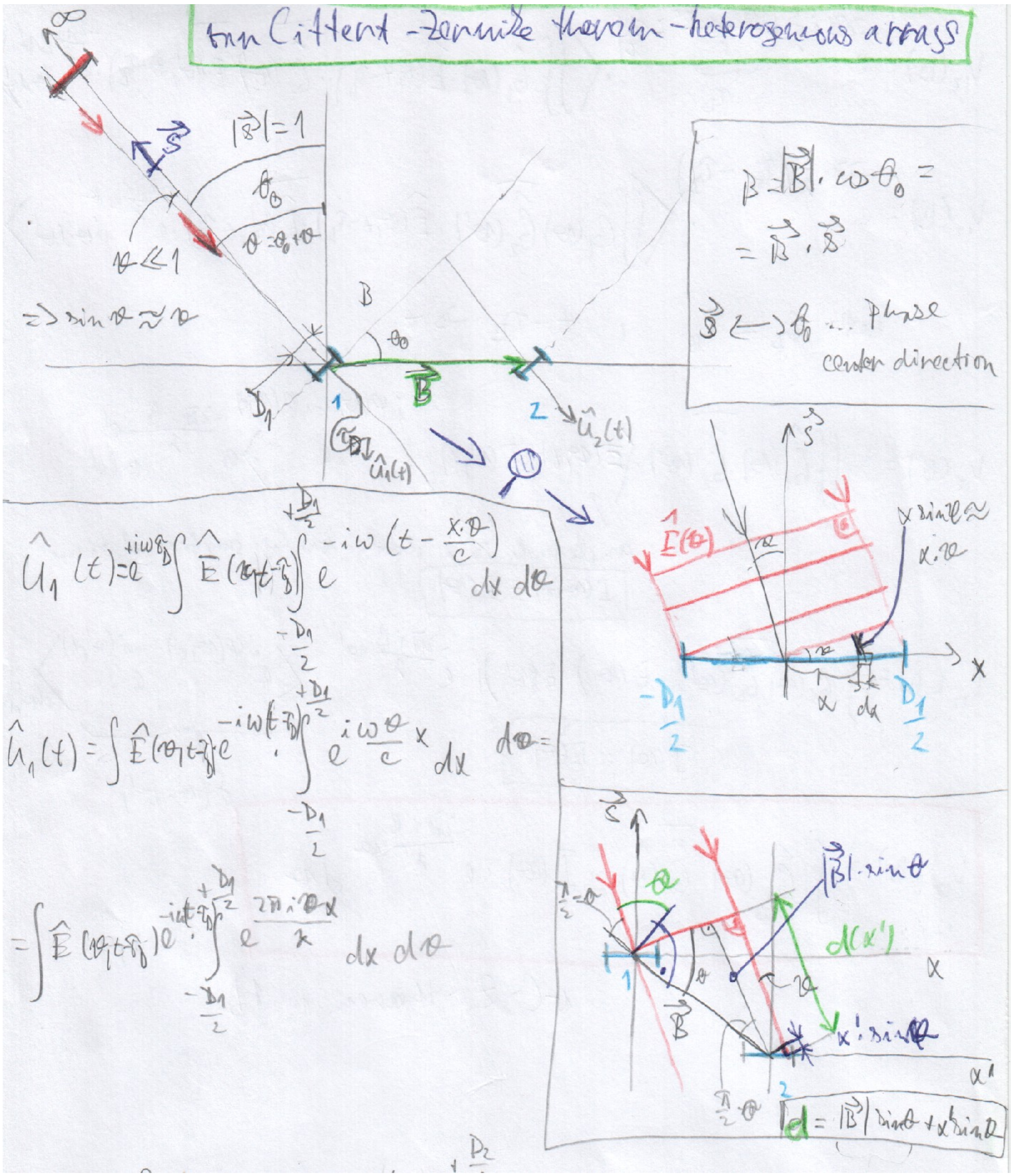


*Fig. 3: Geometry of the two-antenna interferometer.*

source extent, and $\tau_D$ is the usual phase delay introduced by the delay loop in a correlator. We assume that the source extent is not so big, so that the angular distance $\upsilon$ from the phase-reference position $\theta_0$ remains small, and we can thus write $\sin(\upsilon) \approx \upsilon$. The (complex) electric-field amplitude $E(\upsilon,t)$ is assumed to be varying slowly with time, on the time-scale much longer than $1/f$, where $f=c/\lambda$ is the observing frequency. The $K_1$ is an instrument-dependent coefficient for the antenna and receiver 1.

For the *Antenna 2*, the voltage at the same time instant $t$ reads:

$$\hat{U}_2(t)=K_2\int_{source}\hat{E}(\vartheta,t-\tau_B)e^{-i\omega t}\int_{\frac{-D_2}{2}}^{\frac{+D_2}{2}}\exp[\frac{2\pi i}{\lambda}(|\vec{B}|\sin\theta)+x\sin\vartheta]dx\,d\vartheta \quad .$$

Using the relation for $\sin(x+y)$ and, again, taking into account that $\upsilon \ll 1$, one can write

$\sin\theta=\sin(\theta_0+\vartheta)=\sin\theta_0\cos\vartheta+\cos\theta_0\sin\vartheta\approx\sin\theta_0+\cos\theta_0\,\vartheta$ and rewrite the above relation for $U_2(t)$ – with definition of *projected* baseline length (i.e., uv distance) $B\equiv|\boldsymbol{B}|\cos(\theta_0)$ – as

$$\hat{U}_2(t)=K_2\int_{source}\hat{E}(\vartheta,t-\tau_B)e^{-i\omega(t-t_B)}\exp(\frac{2\pi i}{\lambda}B\vartheta)\int_{\frac{-D_2}{2}}^{\frac{+D_2}{2}}\exp(\frac{2\pi i\vartheta x}{\lambda})dx\,d\vartheta \quad .$$

Here, $\tau_B\equiv\frac{|\vec{B}|\sin\theta_0}{c}$ represents the geometrical delay at the *Antenna 2,* which is separated by a baseline of the length $|\boldsymbol{B}|$ from the *Antenna 1,* for waves coming from the phase-reference direction $\theta_0$.

Now, if we define the *electric gain* (a complex-value function, in general) of the antenna $j=1,2$ as

$$G_{E,j}(\vartheta)\equiv K_j\int_{\frac{-D_j}{2}}^{\frac{+D_j}{2}}\exp(\frac{2\pi i\vartheta x}{\lambda})dx=\frac{K_j\lambda}{\pi}\cdot\frac{\sin(\frac{\pi D_j}{\lambda}\vartheta)}{\vartheta} \qquad (1),$$

the (complex) cross-correlation/visibility for the (projected) baseline $B$ $V_{1,2}(B)=\langle U_1(t)\cdot\overline{U_2(t)}\rangle$ (the over-line means complex conjugation and the angle brackets time averaging) can be written as

$$V_{1,2}(B)=\langle U_1(t)\cdot\overline{U_2(t)}\rangle=\int_{source}G_{E,1}(\vartheta)\overline{G_{E,2}}(\vartheta)I(\vartheta)\exp(-\frac{2\pi iB}{\lambda}\vartheta)d\vartheta \qquad (2).$$

Here, $I(\vartheta)=\langle E(\vartheta,t)\cdot\overline{E(\vartheta,t)}\rangle$ is the specific intensity of incoming radiation (i.e., directly proportional to the source brightness). The result above has been achieved using the “standard” assumptions for the van Cittert-Zernike (abbreviated as vC-Z in the following) theorem, i.e., (i) The radiation coming from different part of the source is not correlated (has random phases), (ii) The delay $\tau_D$ introduced in the correlator delay loop is set to equalize the geometrical (baseline) delay $\tau_B$ defined above. If we replace the electric gain $G_E$ by a (differential) antenna area by definition (correct up to some constant real factor for proper units/scaling)

$$A_j(\vartheta)\equiv|G_{E,j}(\vartheta)|^2 \quad ,$$

we can summarize the above result as

$$V_{1,2}(B)=e^{i\Phi_{1,2}}\int_{source}\sqrt{A_1(\vartheta)A_2(\vartheta)}I(\vartheta)\exp(-\frac{2\pi iB}{\lambda}\vartheta)d\vartheta \qquad (3),$$

where the baseline-related constant phase factor in front of the integral may appear if the electrical gains $G_{E1,2}$ have non-zero imaginary parts (ideally they do not). The relation above is basically (1D) generalization of the vC-Z theorem for heterogeneous arrays (i.e., the arrays containing antennas of different sizes), for homogeneous arrays it transforms back to the usual vC-Z formula.

A brief note: This generalization is, actually, not essential for the further discussion about Mars observations. However, I was interested in it as in the *Solar Observing Mode* we use the “kitchen-sink” array, i.e., the combination of the 12m and 7m arrays is not done in post-processing, but it is implemented via hard connection of both arrays into the main (12m) array correlator. Expressing it more explicitly: we have 12-12, 12-7, as well as the 7-7 baselines in the solar science observations. We use *tclean()* with *gridder='mosaicft'*, which should (according to CASA cookbook) ensure the proper combination of the heterogeneous baselines in the Fourier space, but I have to admit that so far it was a kind of a black box for me, I started to think about the theoretical foundations of this combination right inspired by the problem raised by Kazi and Hideo.

Let us now come back to the discussed Mars observations. We want to calculate idealized interference pattern, i.e. , the *flux density* vs. the *uv-distance* for the 12m array. In our 1-D model it can be done by calculating the integral in (2), making its absolute value (modulus), and scaling the result down by the antenna area. For our case $D_1=D_2=D$ being the effective diameter of the 12m antennas. The “antenna area” in our 1-D case means just its “length”, i.e., the diameter $D$. Hence the uv flux density can be calculated as

$$F_{uv}(B)=\frac{|V(B)|}{D} \qquad (4),$$

where $V(B)$ is the integral in (2) for the projected baseline length $B$.

For the Mars disk we use approximation of homogeneous brightness, which in 1-D reads

$$I(\vartheta)=I_0\chi(-M/2,M/2) \qquad (5),$$

where M is the angular diameter of the Mars disk, and $\chi(a,b)$ is the so called characteristic function of the interval $\langle a,b \rangle$, having its value equal to *1* on that interval and *0* outside.

Substituting (1) and (5) into (2), and the result of this substitution finally into (4), we get

$$V(B;D,M)=K^2 I_0 \int_{-M/2}^{+M/2} \left(\frac{\lambda}{\pi\vartheta}\right)^2 \sin^2\left(\frac{\pi D}{\lambda}\vartheta\right) \exp\left(\frac{-2\pi i B}{\lambda}\vartheta\right) d\vartheta \qquad (6),$$

where the $D$ and $M$ after semicolon indicate that the visibility function *parametrically* depends on the (effective) antenna diameter and (effective) Mars-disk size. After integrating by parts and some substitution/re-scaling, and inserting the result into (4), we finally arrive to relation

$$\boxed{F_{uv}(B;D,M)=\frac{K^2 I_0 \lambda}{\pi}\left|-\frac{4\sin^2(m/2)\cos(mb)}{m}+(b-1)Si[(b-1)m]+(b+1)Si[(b+1)m]-2b\,Si(bm)\right|}$$

(7).

Here $Si(x)$ means the (Fresnel) sine integral function, $b\equiv B/D$ is the (projected) baseline length expressed in units of the antenna diameter, and $m\equiv\pi MD/\lambda$ is basically (up to a small numerical factor) the Mars disk (angular) size expressed in units of the antenna primary beam. Relation (7) represents a basis for the further analysis. As can be seen already from (7) and better visible from results of the parametric study (varying *D* and *M*; see further), the i**nterference pattern depends on both antenna diameter *D* and the Mars size *M***.

## Parametric study: varying the antenna diameter(s) and Mars-disk size(s)

We can now play with relation (7) and calculate the interferometric structures for various parameters *D* and *M*. We can also make ratios (like in Fig. 1) between patterns for different set of parameters, just pretending that one set represents the observed and the other one the simulated data. In order to mimic the observed data more realistically we add a small noise to the "observed" flux density.

It is natural to vary the *D* and *M* parameters for the simulated data – such a play is, in fact, the goal of simulations. On the other hand, **we should admit that we do not know exactly even the effective (observation) antenna diameter, say $D_0$, for the 12m array neither the effective (observed) Mars disc size ($M_0$).** The nominal antenna diameter is, of course 12m, on the other hand, for example, for the purposes of *pbcorr()* calculations, the effective antenna diameter has been found to be 10.7m (NB: Wrong setting of this parameter is the root of a known bug, which eventually issued into necessity for re-imaging many older mosaic data in frame of the QA3). I am not an ALMA-technology expert, but I can imagine, that for the purpose of observed interferometric pattern $F_{uv}(B)$ the effective antenna diameter can be yet a bit different. The same holds for the Mars disk size: The homogeneously bright disk of the actual Mars angular size is an idealization, in fact we have different brightness pattern there, which leads to an idea that the homogeneous disk to be used as the best-fit replacement for the actual brightness distribution, can actually easily have different (and unknown) effective size.

In order to perform the parametric study numerically I prepared a set of simple routines in C++ that basically implement the relation (7) plus other above described stuff like adding a small random noise to "observed" data and calculating the sim/"obs" ratio for varying set of parameters *D*, *M*, $D_0$ and $M_0$, and displaying the results via IDL (damned, I still did not find time for learning Python's MatPlotLib and replace/avoid such a way using the stupid IDL ☺). The codes and full set of results can be found – including the description and how-to-use hints (see the *README.pdf* file there) – at the following URL

http://wave.asu.cas.cz/shared/Mars.2019/

In the following I present just a few samples of the results.

## Sample results

First I shall present the result for the case, where the "observation" and simulation parameters match exactly, i.e., $D=D_0$=10.7m (the effective 12-m antenna size as used in *CASA::pbcorr()*) and $M=M_0$=20 arcsecs (as of Hideo's study) – see Fig. 4. The only difference between the two flux densities ("obs" & sim) is just presence of a small additive random noise in the "observed" data. One can see exactly what is expected (cf. Section "Introduction"): The ratio oscillates around the mean value 1.0, the "oscillations" are located around the null points of the interferometric pattern (because of indefinite limit '*0/0*' and the noise present in the "observed" flux density) but the **parts between the adjacent null points are flat, they exhibit no slope, contrary to Fig. 1**.

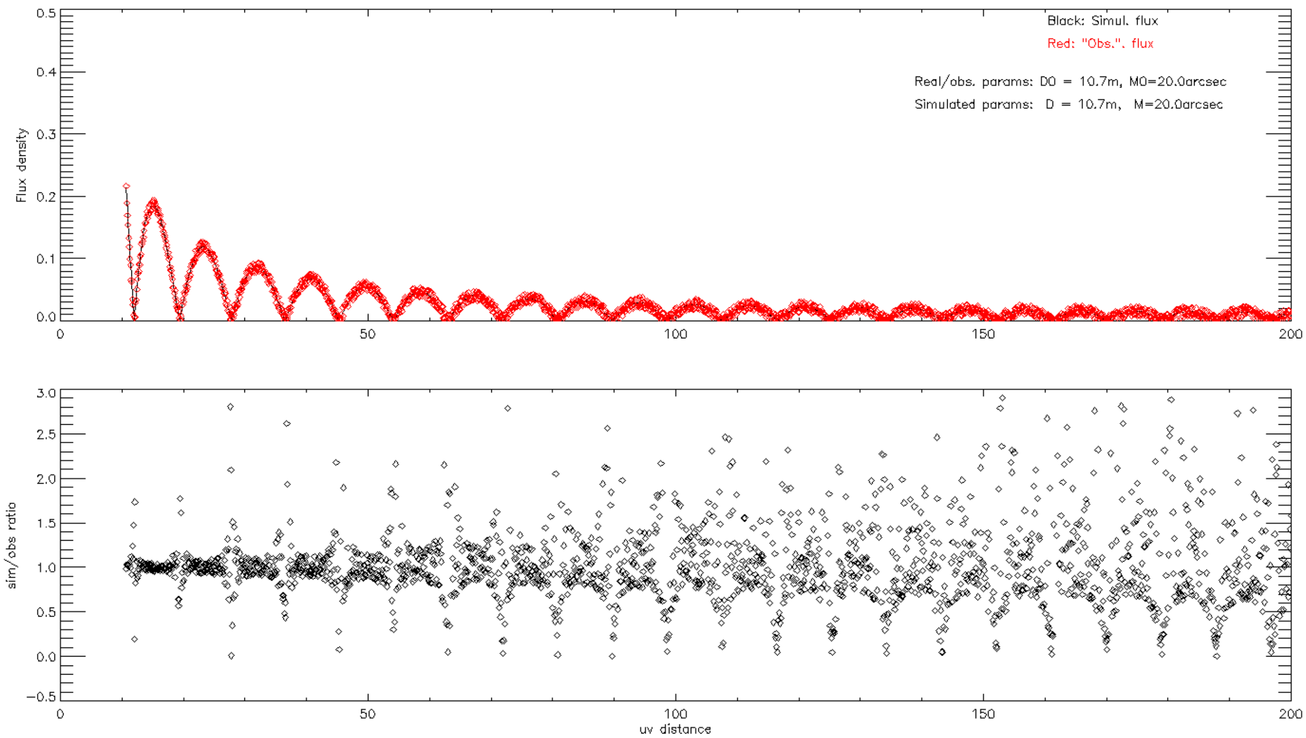


*Fig. 4: Upper panel: Simulated (black line) and "observed" (red diamonds) interferometric pattern (flux density vs uv-distance) for the case where simulated and "observed" antenna & Mars sizes match exactly. Bottom panel: The ratio sim/obs.*

As a next example, I show the data where "observed" and simulated parameters do not match exactly – the "observed" parameters remain the same as before ($D_0$=10.7m, $M_0$=20 arcsecs), while the simulation has used $D$=10.2m and $M$=18.5 arcsec – see Fig. 5 (next page). Even for the relatively small discrepancy between "obs" and sim parameters we get different interferometric patterns. **Namely, the (mean) ratio is not equal to 1.0, and, inherently connected with that, we have the slopes of the sim/obs ratio function between the null points, like in Fig. 1.**

I am aware that there are many issues when comparing our simple 1-D data and the more realistic 2-D study by Hideo. One particular question concerns the relative size of the Mars disc vs. antenna primary beam size. While for standard 2-D parabolic dish the FWHM beamsize reads $\theta_{FWHM} \approx 1.02\,\lambda / D$, in case of the 1-D simplification it is $\theta^{1D}_{FWHM} \approx \frac{2 \cdot 1.392\,\lambda}{\pi D} \approx 0.886\,\lambda / D$. (NB: The numerical factor 1.392 has been found by solving the transcendental equation $\sin^2(x)/x^2 = 1/2$, which implies from relation (1) – see the toy code *fwhm.cc* in the *parametric_study/* subdir at the above referenced URL). Hence, the Mars disk size as expressed in the antenna FWHM beam size is different for the 1-D case if we use the "2D - study" parameters $D_0$=10.7m and $M_0$=20 arcsecs. In the 2D study by Hideo (and in observed reality) the size of the Mars disk expressed in PB size is ~20"/26"=0.77. In order to fit our 1D simulation to this value I have mangled both the "real" antenna diameter and the "real" Mars size to values $D_0$=8m and $M_0$=15arcsecs and have made a second set of parametric study for those "observation/real" set of parameters: That is why there are two sub-directories in the *fig/* directory – one to fit the real antenna and Mars sizes to the 2D study and one to match rather relative Mars disk size expressed in terms of the primary-beam FWHM between 1D and 2D cases. One example from the latter set of the parametric-study results is in Fig. 6 (page 9). However, qualitatively there is no significant difference.

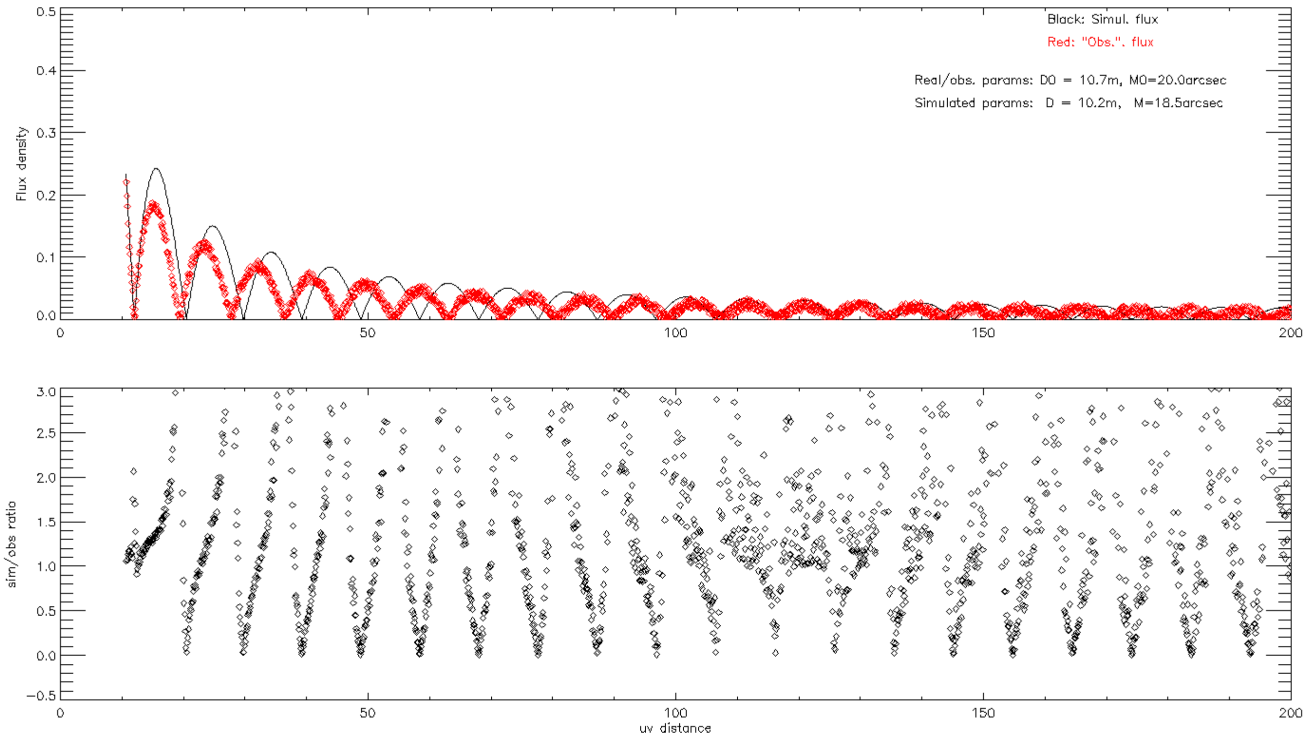


*Fig. 5: Like in Fig. 4, but the sim and obs parameters mismatch. This has two mutually related consequences: The mean of ratio is not 1.0 and the slopes in ratio (cf. Fig. 1) between the null points are prominent.*

## Conclusions

This study aims at contributing to the discussion, whether the Mars observation by 12m ALMA array has trustworthy flux densities or whether these fluxes need to be up-scaled by some factor (about ~1.3). I did a simple 1D analytical calculations in order to find, which parameters control the shape of interference pattern (= flux density expressed as a function of projected baseline length [a.k.a. the uv-distance]). Being just 1D, the analysis is certainly a simplification that can not reproduce the real 2D situation in Hideo's analysis quantitatively. On the other hand, qualitatively, the main features are reproduced well and as the resulting relation (7) for the flux density is an analytical expression, we have full control over parameters that change the shape of interferometric pattern. Namely, it is clear that **both the antenna diameter *D* and the Mars disk size *M* influence the pattern**, including the positions of its null points.

The results found from the parametric study based on the relation (7) can be summarized as follows:

- Ratio between interferometric patterns with different controlling parameters *D* and *M* is a quite sensitive indicator – even small differences in controlling parameters lead to a large change of the pattern from the case of the perfect match displayed in Fig. 4.
- **Mismatch between the "observation" and "simulation" parameters *D* and *M* expresses itself in the sim/obs ration by** (i) departure of the ratio mean form 1.0, and (ii) **clearly visible slopes between the null points of the interferometric pattern.**

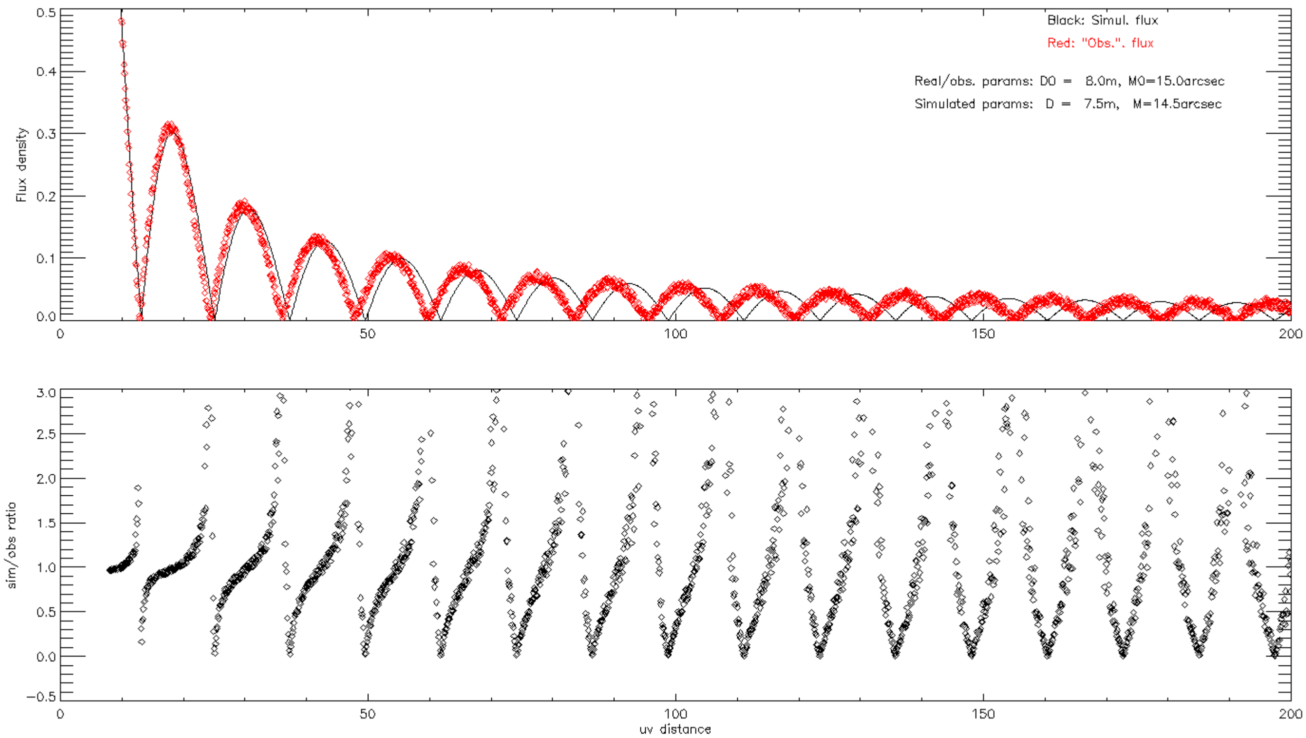


*Fig. 6: Like in Figs. 4 and 5, the "real/observation" antenna diameter and Mars angular size were both mangled to fit the ratio of Mars angular size to primary-beam FWHM from the 2D Hideo's study.*

- **Because the data in Hideo's study clearly show similar features – namely the strong slopes, I would argue that the effective antenna size and/or effective Mars-disk size as used in the simulation does not match the same parameters in reality/observations.** This mismatch might be caused by reasons that I have discussed above – let me briefly repeat: (i) the effective antenna diameter for the purpose of measuring the interferometric pattern might be different from both the nominal value of 12m and the effective diameter of 10.7m as used in CASA::pbcorr(), (ii) the size of the effective homogeneously-bright Mars disk, which we take as a replacement for the actual inhomogeneous Mars brightness distribution, might be different from its nominal "optical" angular diameter. Or, even, such a replacement is inappropriate and the sim/obs ratio is very sensitive to the exact Mars brightness distribution, as Hideo already pointed out.

To sum up, I do not see the **mismatch between simulated and observed 12m data in the Hideo's study** as critical, because – at least to some extent – they **might be explained by the mismatch between effective antenna and/or Mars sizes used in the simulations and the reality**. Hence, the **observed 12m flux is not precluded to be wrong** and I would tend rather to believe it.

In Chlum and Ondřejov, November 14th, 2019